View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Environmental Sciences 17 (2013) 317 – 326
Procedia Environmental Sciences

The 3rd International Conference on Sustainable Future for Human Security
SUSTAIN 2012

Flood prediction using transfer function model of rainfall and water discharge approach in Katulampa dam

Yulianti Hasanah[a,*], Marizsa Herlina[a], Hilda Zaikarina[a]

[a]*Department of Statistics Bogor Agricultural University, Bogor 16680, Indonesia*

Abstract

Flood in Indonesia has become a usual event of disasters every year, especially in Jakarta. Losses due to floods in Indonesia are not only a small value, so therefore it is required to have an early warning before flood occurs. Floods can be identified by observing the river water discharge. Transfer function model is used because it is assumed that rainfall is having an influence of water discharge. The transfer function model between Ciliwung River water discharge and rainfall that located at Katulampa Dam, the Dam of Ciliwung River, East Bogor, this model could be a solution for early warning of floods. The results of the research shows that the transfer function model which has been obtained can explain the relationship between the water discharge and the rainfall for two months previously. The results of the forecast using the transfer function approach the actual data closely for the first two months after the final data input. Transfer function model is a better forecast model than ARIMA model of water discharge.

© 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of SUSTAIN conference's committee and supported by Kyoto University; (OPIR), (GCOE-ES), (GCOE-HSE), (CSEAS), (RISH), (GCOE-ARS) and (GSS) as co-hosts.

Keywords: transfer function model; flood; water discharge forecast; rainfall

1. Introduction

Indonesia is a country that has a very high potential for disaster. The potential range includes floods and landslides, especially in the rainy season. Floods are a regular phenomenon encountered in some areas that causes huge losses, such as floods in February 2007 in the Jakarta area, The capital city of Indonesia, for five days reached 8.6 trillion losses, equivalent to 48% in 2006 Capital Budget with 60 victims of the 263.416 displaced people.

* Corresponding author. Tel.: +6287770843774
E-mail address: yuliantihasanah.stk@gmail.com.

1878-0296 © 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of SUSTAIN conference's committee and supported by Kyoto University; (OPIR), (GCOE-ES), (GCOE-HSE), (CSEAS), (RISH), (GCOE-ARS) and (GSS) as co-hosts.
doi:10.1016/j.proenv.2013.02.044

Based on those statistics, it can be said that the losses resulting from flood is quite large, therefore it is required early notification before a disaster occurs to minimize the losses, one of the solution to solve this problem is using flood forecasting. Jakarta is a city with a high potential of flooding, it is shown that each year floods always occur in Jakarta. And it is also assumed that floods in Jakarta are also the result of rain water that flows from Bogor as the upstream of the Ciliwung river, precisely in Katulampa, East Bogor. Bogor is a city with a high rainfall. The surface of Bogor which is higher than Jakarta causes the rain water that fell on Bogor affected the Ciliwung River that irrigates Jakarta on the lower surface, resulting in overflowing of the Ciliwung River in Jakarta that led to the area around the river floods. So this study took place in Ciliwung River Dam, Katulampa, East Bogor.

Water discharge is the volume of water flowing through a river cross section per unit time, in unit's m^3/second. Extreme water discharge indicates that the river will be prone to overflow, so water discharge may become one of flood's indicators. According to researchers, because of the correlation between rainfall and water discharge, we suggest a model that can forecast the water discharge from the rainfall and water discharge, the model is the transfer function with a rainfall as the input series and water discharge as the output series. This study traced the forecasting model of water discharge of Ciliwung River from Katulampa rainfall and water discharge of Ciliwung River.

2. Methodology

Transfer Function Model is a model that combines time series approach with a causal approach. The time series x_t influences the time series y_t through a transfer function which distributes the impact of x_t through some period in the future. The resulting model is called the transfer function model which connects the output series (y_t), the input series (x_t), and noise (n_t) [5]. The correlation between X and Y (transfer function) is also called the cross-correlation. The rainfall and water discharge transfer function model procedure includes the following stages:

2.1. Exploration of the rainfall data and the water discharge data

The data that we used are the monthly rainfall data and monthly water discharge data from January 1996 until December 2011. The Rainfall data and water discharge data are explored using a descriptive analysis.

2.2. Prepare output series and input series (Data Stationarity)

$$Zt = at + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots.. \quad (1)$$

a_t is a white noise which form rows of random variables which is independent and have identical distribution. General model of time series models include a more specific, such as the mixture process between the two models which is Autoregressive-Moving Average commonly called ARMA models. Generate both of the time series plot for Rainfall data as an input series and water discharge as output series. Augmented Dicky-Fuller test is used to test the stationary of the data.

2.3. ARIMA model identification for all variables

Autoregressive Integrated Moving Average (ARIMA): a mixture of regression models of order p and moving average of order q. General model of time series data is ARIMA (*p,* d*,* q*)* with the general model*:*

$$\O_p(B)\, \nabla^d Z_t = \theta_q(B)\, a_t \quad (2)$$

where:
Ø = autoregressive parameter
Θ = moving average parameter
a_t = random error at period t which is assumed to be normally distributed and stochastic independent
∇^d = operator of differentiation with d degree of differentiation.
$Ø_p(B) = (1-Ø_1 B - ... - Ø_p B^p)$ is AR characteristic polinomial.
$\Theta_q(B) = (1 - \theta_1 B - ... - \theta_q B^q)$ is MA characteristic polinomial.

If the specified value of q = 0, the model will become an autoregressive model of order p or AR (p). Conversely if it is determined that the p = 0 the model will become a moving average model of order q which is abbreviated MA (q).

2.3.1. Criteria for Model Selection

Schwarz's Bayesian Criterion (SBC) or also called Bayesian Information Criterion (BIC) is used as a criterion for selecting models. SBC is a model criterion selection based on maximum likelihood function. The best model is the model with minimum SBC value. Akaike Information Criterion (AIC) tends to choose a model with more parameters of the SBC. For large data SBC is better and more consistent. After doing the forecasting, the accuracy of forecasting can be searched by calculating the Mean Absolute Percentage Error (MAPE),with the following formula:

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \frac{x_t - f_t}{x_t}}{n} \times 100 \quad (3)$$

Where x_t is an observation at t-period and ft is a forecast at t-period. The smaller the value of MAPE indicates that the data forecasting results closer to the actual value.

2.4. Prewhitening rainfall input series and water discharge output series

Prewhitening is a series of transformation processes which are correlated to the behavior of white noise which are not correlate. This process uses ARIMA models for the input series. Therefore, before the prewhitening process, the ARIMA models for x_t are constructed. Because the input series has undergone prewhitening (αt) the model becomes:

$$\alpha_t = Ø_p (B)\theta_{q-1} (B)x_t \quad (4)$$

Transfer function is a x_t mapping process on y_t. So that when a prewhitening process of xt is applied, then the same transformation must be applied for yt to maintain the integrity of functional connections. So that the output series that has been transformed (βt) is:

$$\beta_t = Ø_p (B)\theta_{q-1} (B)y_t \quad (5)$$

2.5. Calculate the cross-correlation between the input and output series.

The correlation functions between αt and βt at the *k*-th lag is:

$$\hat{\rho}_{\alpha\beta}(k) = \frac{\hat{\gamma}_{\alpha\beta}(k)}{s_\alpha s_\beta},\ k = 0, \pm 1, \pm 2, \ldots \qquad (6)$$

where:
$\hat{\rho}_{\alpha\beta}(k)$ = Cross-correlation between α_t dan β_t at the *k*-th *lag*
Covarian between
$\hat{\gamma}_{\alpha\beta}(k)$ = Covarian between αt dan βt at the *k*-th *lag*.
s_α = standar deviation of α_t series
s_β = standar deviation β_t series

2.6. Initial Identification of transfer function model

The constants b, r, and s is determined by the plot pattern of cross-correlation function between αt and βt. How to determine the value of b, r, and s is (1) cross-correlation significantly different from zero for the first time in the lag-b. (2) For s is seen from the subsequent lag that has a clear pattern or how long x influences y after the first significant lag.(3) The r value indicates how long the output series (y_t) are associated with the previous value of the output series itself. Value of r can be seen from the plot of autocorrelation *yt* or determined by the plot pattern of lag (b + s), if it has an exponential plot pattern then the r = 1 and if it has a sine wave plot pattern then the r = 2 [1].

2.7. Initial parameter estimation δ and ω

Initial estimates of the transfer function parameters are $\hat{\delta}$=(δ_1, δ_2,…, δ_r) and $\hat{\omega}$=(ω_0, ω_1,…, ω_s) and those are sought using the following equation:

$$\begin{aligned} &Vj = 0, j < b \\ &Vj = \delta 1Vj\text{-}1 + \delta 2Vj\text{-}2 + \ldots + \delta rVj\text{-}r + \omega 0, j = b \\ &Vj = \delta 1Vj\text{-}1 + \delta 2Vj\text{-}2 + \ldots + \delta rVj\text{-}r - \omega j\text{-}b, j = b+1, \ldots, b+s \\ &Vj = \delta 1Vj\text{-}1 + \delta 2Vj\text{-}2 + \ldots + \delta rVj\text{-}r + \omega 0, j > b + s \end{aligned} \qquad (7)$$

Where

$$\widehat{V_k} = \frac{r_{\alpha\beta}(k)s_\beta}{s_\alpha} \qquad (8)$$

This initial estimate is used as the initial value at the end of the nonlinear estimation algorithm and to estimate the residual series.

2.8. Initial identification of residual model

Identifying the model with a residual series that has been obtained after the initial parameter estimates to examine whether the model is good or not and it is examined in the same manner as the previous model identification.

2.9. Determine the combination of transfer function model

2.9.1. Final parameter estimation of transfer function model

Initial parameter estimation is the initial value of the nonlinear least squares estimation algorithms to form the final estimate for the model parameters which are carried out iteratively. The process is repeated until it is convergent. Nonlinear least squares estimation is to estimate the parameters obtained by minimizing:

$$S(\delta, \omega, \phi, \theta | b) = \sum_{t=t_0}^{n} a_t^2 \qquad (9)$$

2.9.2. Model diagnostic of transfer function model

Goodness of fit checks conducted by looking at the behavior of the residual (*at*) and sample cross-correlation (SCC) between a_t and α_t (residual and input). Randomness of the residual and the lack of significantly different from zero SCC values are indicates the model is appropriate. Box-Pierce Q statistic test can be applied to test the independency of the residual and the absence of correlation between the input and the residual.

2.10. Forecast the monthly water discharges at Katulampa Dam using the best model

Forecast is calculated using the equation:

$$\delta_r(B)\O_p(B)y_t{=}\O_p(B)\omega_s(B)X_{t-b}+ \delta_r(B)\theta_q(B)a_t \qquad (10)$$

by entering the parameter values of the transfer function and the value of input and output series are obtained from the previous steps.

2.11. Compare the forecast using the transfer function and ARIMA model

At this stage the best transfer function model will be compared with the output series ARIMA models and selected the model with the smallest MAPE as the best model.

3. Result

Rainfall in Katulampa highly fluctuated each month and every year. The highest rainfall in the month of November 2004 was 816 mm and it reached the lowest point 4 mm in July 2003. In Ciliwung river water discharge reached its highest in February 2007 as much as 58.517 ltr / sec and the lowest point reached 1.26 liters / sec occurred in July 2006. P-values based on the correlation between rainfall and water discharge is 0.434, shows that rainfall associated with the water discharge directly.

3.1. Preparing Input Series and Output Series (Data Stasionarity)

The stationarity of the data is tested using the Augmented Dickey-Fuller test. The test results obtained show that the p value was less than 0.05, this shows that the data has been stationary. It could be seen as follows, testing the original data of the x_t and y_t data shows that the data are stationary.

Table 1. the value of Augmented Dickey-Fuller test statistic

	t-Statistic	Prob.*
Xt	-9.164978	0.0000
Yt	-7.617367	0.0000

3.2. ARIMA Model Identification of Rainfall

ACF and PACF Plot from the stationary input series x_t

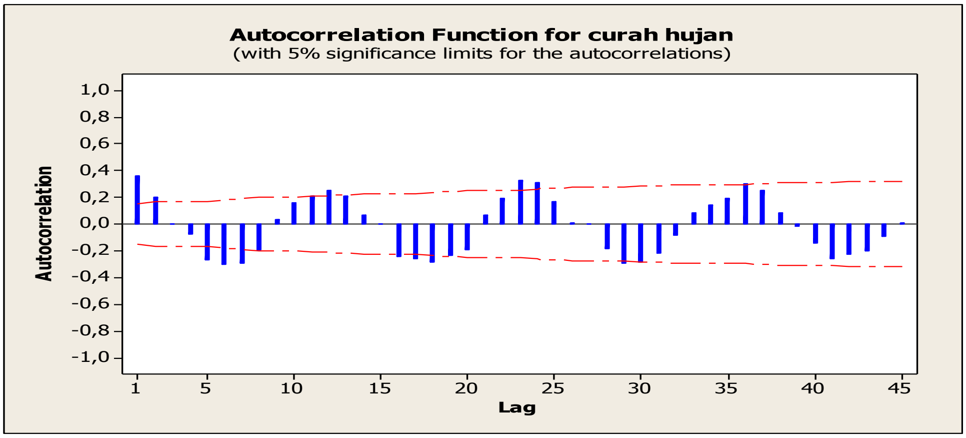

Figure 1. ACF Plot of Input series x_t

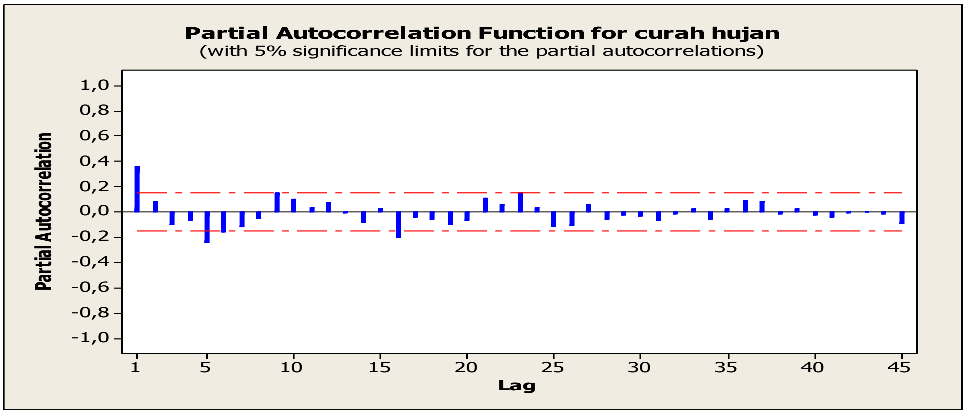

Figure 2. PACF Plot of Input series x_t

Table 2. The value of AIC and SBC ARIMA models of input series x_t

Model	AIC	SBC
AR* (1)	2350.22	2356.606
MA* (2)	2350.395	2359.974
ARIMA (1,0,2)	2352.013	2364.785
ARIMA* (2,0,1)	**2345.428**	**2358.2**
ARIMA (2,0,2)	2353.281	2369.245

Note : (*) Parameter (θ) is significant

Table 2 shows that the model ARIMA (2,0,1) is the best model because it has the smallest value of AIC and SBC as compared to other ARIMA models and all of the parameters are significant. So ARIMA models of rainfall obtained are:

$$X_t = 1{,}33577\,X_{t-1} + 0{,}38418\,X_{t-2} - \alpha_{t-1} + \alpha_t + 333{,}29754 \tag{11}$$

3.3. ARIMA Model Identification of Water Discharge

Here is the ACF and PACF plot output series y_t which has been stationarized.

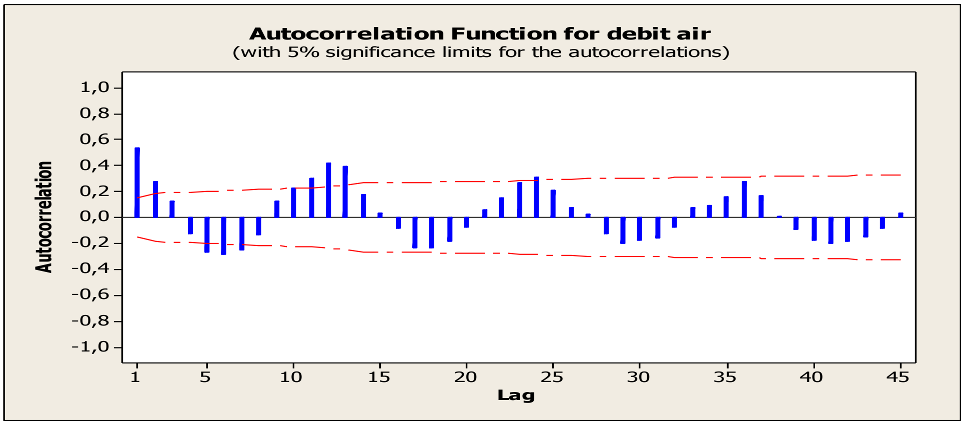

Figure 3. ACF plot of output series y_t

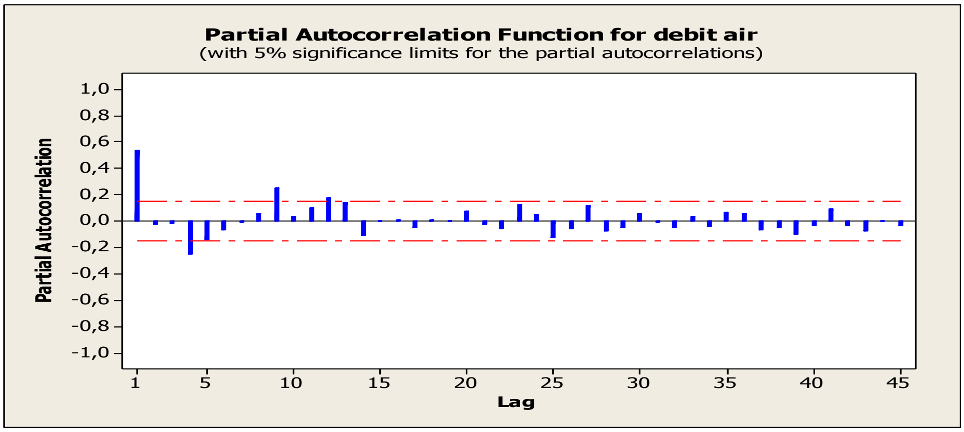

Figure 4. PACF plot of output series y_t

Table 3. The value of AIC and SBC ARIMA models of input series y_t

Model	AIC	SBC
AR* (1)	1333.551	1339.937
MA* (2)	1340.629	1350.208
ARIMA (1,0,2)	1337.484	1350.256
ARIMA* (2,0,1)	1334.567	1347.339
ARIMA* (2,0,2)	**1323.215**	**1339.18**

Table 3 shows that the model ARIMA (2,0,2) is the best model because it has the smallest value of AIC and SBC as compared to other ARIMA models and all of the parameters are significant. In addition, the Box-Pierce test shows that the autocorrelation value of the residuals is not zero ($\alpha = 0.05$) for all lags. This means that the residual is not correlated with each other. So ARIMA models of water discharges obtained are:

$$Y_t = 1{,}33336\, Y_{t-1} + 0{,}67383\, Y_{t-2} - 1{,}64024\, \beta_{t-1} + 0{,}90776\, \beta_{t-2} + \beta_t + 11{,}92788 \tag{12}$$

Input Series (Rainfall) and Output Series (Water Discharge) Prewhitening

Prewhitening stage is based on the ARIMA models for the rainfall. In this phase, the white noise model elements are used. Thus the prewhitening models for the input series x_t is:

$$\alpha_t = X_t - 1.33577\, X_{t-1} - 0.38418 X_{t-2} + \alpha_{t-1} - 333{,}29754 \tag{13}$$

Prewhitening process in the output series also follows the prewhitening model for the input series. So as to obtain the white noise β_t of the output series y_t as the following:

$$\beta_t = Y_t - 1.33577\, Y_{t-1} - 0.38418 Y_{t-2} + \beta_{t-1} - 11{,}92788 \tag{14}$$

3.4. Crosscorrelations of the input and output series

Cross-correlation shows a relationship between rainfall and water discharges. The pattern of Cross-correlation will be used for identification of transfer function model (b, r, s). The results of the cross-correlation between α_t and β_t that has been calculated are b = 0, s = 1 and r = 2.

3.5. Initial identification of Transfer Function Model

Initial identification of the model is done by looking at the plot pattern of cross-correlation between α_t and β_t. Based on the information above, the initial identification of transfer function model has a value of b = 0, s = 1, and r = 2. An examination of the model have to be conducted in order to to get the best models. General form of the transfer function model is:

$$y_t = \left(\frac{\omega_0 - \omega_1 B - \cdots - \omega_s B^s}{1 - \delta_1 B - \cdots - \delta_r B^r}\right) x_t + n_t \tag{15}$$

So the initial identification transfer function model is as follow:

$$y_t = \left(\frac{0{,}02291 + 0{,}01793B}{1 + 0{,}12484B - 0{,}43748B^2}\right) x_t + n_t \tag{16}$$

3.6. Identification Model of Residual

The model obtained from the initial identification of transfer function models, namely:

$$\mathrm{y_t} = \left(\frac{0{,}02291 + 0{,}01793\mathrm{B}}{1 + 0{,}12484\mathrm{B} - 0{,}43748\mathrm{B}^2}\right) \mathrm{x_t} + \mathrm{n_t} \tag{17}$$

So as to obtain the value of n^t:

$$n_t = y_t - \left(\frac{0{,}02291+0{,}01793B}{1+0{,}12484B-0{,}43748B^2}\right) x_t \tag{18}$$

Initial identification of transfer function model produce plots of residual ACF and residual PACF. ACF plot shows that it is significant at lag 2 and the PACF plot shows that it is significant at lag 1. So that initial identification of residual model is ARIMA (1,0,0).

$$n_t = \left(\frac{1}{1-\Phi_1 B}\right) a_t \tag{19}$$

Table 4. Table of initial identification of transfer function model

no	b,r,s value	parameter	t-value	AIC	SBC
1	(0,2,1)	ω0 ω1 δ1 δ2	5.66 -0.42 -0.06 0.36	1190.005	1209.914
2	**(1,2,1)**	**ω0** **ω1** **ω2** **δ1**	**0,68** **-1,82** **95,07** **56,66**	**1186.845**	**1205.199**
3	(0,2,2)	ω0 ω1 ω2 δ1 δ2	4.31 3.59 -3.44 33.95 -20.49	1192.081	1214.172
4	(1,2,2)	ω0 ω1 ω2 δ1 δ2	3.21 0.86 -2.47 7.81 -7.43	1197.609	1219.665

Based on Table 4, it can be seen that the number 2 model indicates that all of the parameters are non-zero coefficients. The smallest AIC and SBC values contained in the model values of b = 1, r = 2, and s =1 with a standard error of 8.588943.

3.7. Final Parameters Identification of Transfer Function Model

Final identification of the transfer function model is done by combining the initial model with the residual. Based on the residual autocorrelation value, the residual value of the transfer function model are independent because the autocorrelation and cross-correlation between the input series and the residual is zero (α = 0.05). In consideration of parameter test, autocorrelation of residual, and correlation between input series and residual, it was determined that the final transfer function model is:

$$y_t = \left(\frac{\omega_0-\omega_1 B^1}{1-\delta_0 B-\delta_2 B^1}\right) x_{t-b} + \left(\frac{1}{1-\Phi_1 B}\right) a_t \tag{20}$$

$$y_t = \left(\frac{0{,}00068+0{,}0019B}{1-1{,}66516B+0{,}98309B^2}\right) x_{t-1} + \left(\frac{1}{1-0{,}4108B}\right) a_t \tag{21}$$

$$yt = 1{,}66612x_{t-1} - 0{,}98387x_{t-2} + y_{t-1} + 0{,}4108y_{t-2} \tag{22}$$

This transfer function model means that water discharges influenced by rainfall intensity for one month and two months previously and also the water discharge for one month and two months previously.

3.8. Forecasting

To find out about the accuracy and effectiveness of water discharge estimation based on the models obtained, a model validation is needed. The concept of model validation is the comparison between the actual data and the data obtained from the forecasting model that has been produced. Comparison of transfer function models and ARIMA models with the actual data can be seen in Table 5. MAPE value of forecasting using the transfer function model is 15.23, while the ARIMA models for 51.34. Based on table 5, note that the forecasting model of the transfer function is closer to the actual data compared to the ARIMA models of water discharge.

Table 5. Table of Water Discharge Forecast and MAPE Value using Transfer Function Model and ARIMA Model

Month	Forecast		Actual
	Transfer	ARIMA	
January	22.529	25.66	20.166
February	26.521	23.115	32.642
MAPE	15.23%	56.43%	

4. **Conclusion**

The results of the research shows that the transfer function model which has been obtained can explain the relationship between the water discharge and the rainfall two months previously. The results of the forecast using the transfer function approach the actual data for the first two months and this model is a better forecast model than ARIMA model of water discharge because the MAPE value of transfer function model is less than MAPE value of water discharge ARIMA model. For further research, it would be more accurate if the model included more variables other than rainfall to predict the water discharge.

Acknowledgements

Thank you for Bogor Agricultural University, Faculty of Mathematics and Natural Science, Department of Statistics, Yenny Anggraini, M.Si as lecturer, BMKG Station Dramaga, PSDA Paledang, PT. Harrosa Mitra Jaya and Detroit Autoworks for supporting this research and publication for supporting this research and publication.

References

[1] Bowerman BL, Richard T.O'Connell.1993. *Forecasting and Times Series :an applied approach.* 3rd edition. California: Wadsworth.
[2] Box GEP, GM Jenkins, GC Reinsel. 2008. Times Series Analysis
[3] Cryer, JD. 1986. *Time Series Analysis*. Boston: Duxbury Press.
[4] Indra Prasetyo, Erwin. 2009. *Analisis Hubungan Curah Hujan dan Produksi Kelapa Sawit dengan Model Fungsi Transfer*. Skripsi. Bogor: Institut Pertanian Bogor.
[5] Makridaskis S, SC Wheelwright, VE Megee.1983..*Forecasting: Methods and Applications*. 2nd edition. New York:John Wiley and Sons.
[6] Wei, WWS. 1990. *Time Series AnalysisUnivariate and Multivariate Methods*.Canada: Addison-Wesley.